Can Data Representation and Interface Demands be reconciled? Approach in ORCA

A.M. van Ginneken, M. de Wilde, E.M. van Mulligen, H.Stam
Department of Medical Informatics
Erasmus University, Rotterdam, The Netherlands

Research in the domain of computer-based patient records has always faced the conflicting demands of efficiency for the practicing physician and suitability of the record contents for data analysis in view of decision support, research, and quality assessment. Interface and contents pose different demands on the data model underlying the record. The challenge is to combine the most suitable model for data representation with the interface that best fits the clinical setting. ORCA (Open Record for CAre) provides a solution by making the distinction between domain dependent and domain independent data and letting domain dependence be decisive for the choice of model. Interactive definition of custom-views provides interface flexibility for domain dependent data. Views on domain independent data need not cope with the limitations of multiple table views in relational DBMSs. A standard set of single table queries can support recording of domain independent data, irrespective of the clinical setting.

INTRODUCTION

The rapid increase in medical knowledge and technology has led to a complex health care system, which is characterized by superspecialization and fragmentation of patient data. In this environment, sharing of data, decision-support, quality assessment, and research become ever more important [1-4]. It is widely understood that the recording of data in free text permits the most freedom in working style, but it is not a suitable format for data-analysis and interpretation [5].

The collection of structured, codable data has received the attention of many researchers for almost a quarter century. Part of the research focused on natural language processing (NLP), which attempts to extract structured data from free text [6-7]. Because of its little impact on a clinician's working style, it was believed to be more easily accepted than structured data entry (SDE), which requires the recording of findings in a structured format at the time they are collected [8]. Several studies have demonstrated that SDE has a far greater potential for the quality and completeness of the data being collected [9-11]. Therefore, continuing effort is directed at making SDE feasible within the time-pressured environment of the practicing clinician [12-17].

SDE can be supported by a fixed or dynamic interface strategy. The first refers to forms with a predefined layout and content, whereas the latter denotes an interface where descriptive options are dynamically generated during interaction with the user. Which type of interface is most efficient depends on the clinical setting [18]. On the other hand, a distinction between domain dependent and domain independent data is important for the choice of a direct or indirect model to support SDE [19].

Designing a patient record that combines the strengths of both models is not self-evident. The preferred interface and the type of data involved are determined by two different aspects of the clinical situation: the scope and the content, respectively. To complete the framework, we need insight in the relationships between the clinical setting and the nature of the data to be recorded.

In this paper, we will discuss the importance of scope and content for the choice of data model and address the challenges we encountered in our attempts to apply the most suitable strategies for the clinical setting involved. Subsequently, we will present our solutions to those challenges.

THE ROLE OF INTERFACE, CONTENT, AND DATAMODEL

Clinical Setting and interface demands

In view of consequences for data entry, there are two main types of clinical setting: 1) the collection of a well-defined data set in the context of a research or clinical protocol, and 2) situations where physicians are confronted with a large variety of abnormalities, such as in primary care and internal medicine. In the first setting, the scope is small and data entry is predictable because of its mandatory nature. In the second setting, the scope is large and data entry mainly unpredictable, because of many accidental findings.

Predictable data entry is best supported with fixed forms, that contain all items to be recorded in a wellorganized layout, tailored to the order in which the data are collected. Fixed forms are not the first choice for unpredictable data entry. When the domain is large and the number of findings to be entered is relatively small, data entry is best served with a dynamic interface that offers descriptive options in the context of the findings or topics at hand.

Data models and data categories

Support of SDE requires knowledge about the data items to be entered. This knowledge involves the descriptors of medical concepts: e.g. a blood pressure is described by a systolic and diastolic pressure, whereas a drug prescription is described by the drug name, the dosage, and the frequency of intake. Which model is most suitable to represent and apply that knowledge for the support of SDE strongly depends on how variable the descriptors of a concept may be. In view of the descriptors involved, we can distinguish two main categories of patient data: categories where the concept descriptors are domain independent versus categories where the concept descriptors are domain dependent, i.e. vary per domain. Typical examples of domain independent data categories are drug prescriptions and laboratory data. Irrespective of the type of specialists requesting the test, a laboratory test result is described by the name of the test, the value, and the units. In contrast, data recorded during history taking and physical examination vary greatly per specialty in scope, focus, and detail. One fixed set of attributes to record the results of history taking or physical examination will never satisfy all specialists.

The choice of model. Direct models are characterized by a direct mapping between the entry fields on the screen and the attributes in a relational database. Examples are COSTAR, TMR, Regenstrief record, BAZIS [8,20]. Direct models have the advantage of being supported by a variety of software packages and retrieval is straightforward with standard SQL queries. Direct models are quickly developed for well-defined data sets of limited scope. However, adjustments are labor-intensive in the sense that every change to the data set to be recorded requires modifications of interface, database, and queries. When scaling up, adjustment, expansion, and documentation efforts will grow exponentially. This relatively rigidity, however, is of little importance in domains where few changes need to be made. Hence, a direct model is a natural choice to support SDE for domain independent data. Indirect models are often used in applications for knowledge-driven data entry, where descriptive options are dynamically created on the basis of a controlled vocabulary and user input [12,15,17,21,22]. The controlled vocabulary defines which terms can be used and how they may be combined to form medically meaningful descriptions. The actual patient data are instantiations of the concepts in the controlled vocabulary. The term 'indirect' pertains to the fact that the format in which the patient data are represented does not permit a direct mapping with the options for data entry. In ORCA, the instantiated patient data are numbered nodes in a tree, that need to be mapped to the thesaurus of the controlled vocabulary to become meaningful. The indirection makes it possible to separate database structure and content, which is the key to the flexibility of these models. Changing the contents does not require changes to the database structure and the interface. Retrieval, however, requires a dedicated algorithm, which makes exchange of data with other applications more complex. Because of their flexibility, indirect models are the preferred choice for the recording of domain dependent data.

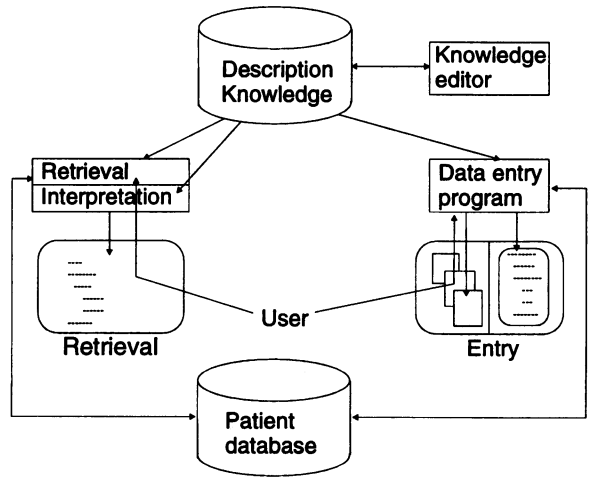

Fig. 1. The indirect model in ORCA. The concepts in the knowledge base are needed to interpret the patient data.

Clinical setting and data categories

Based on the insights gained, the design of a CPR that combines the strengths of the two models may seem straightforward: a direct model with fixed forms to support the entry of small well-defined data sets and an indirect model with dynamic forms for the entry of accidental findings. This approach, however, only takes into account the best fit between model type and interface strategy. Hence, it assumes that the model that best supports the clinical setting corresponds with the type of data to be entered. In other words, small well-defined data sets would correspond with domain independent data, and accidental findings with domain dependent data. In practice, there is no such relationship between clinical setting and type of data to be entered. On the contrary, in a highly specialized setting, the data set is small and well-defined, but very specific for the domain.

The question arises whether it is correct to conclude that the entry of domain-dependent data is best served with an indirect model. Is not the scope of the data to be entered, and hence, the preferred interface strategy decisive for the choice of model? There are several reasons why the use of a direct model for domain dependent, but well-defined data sets is not desirable.

Multiple representations of single concepts. When a direct model is used in an application that tailors its records to specific needs, overlapping items such as blood pressure, pulse rate, and weight, may be represented by different attributes in different tables. Queries such as 'retrieve all blood pressure measurements of a selected patient', become complex and must include all attributes that represent a blood pressure. Beside the complexity of such queries, their reliability is subject to the programmer's understanding of the physician's needs, and how well the database documentation is kept up-to-date. Many existing databases do not permit straightforward queries, because the relationship between semantics and structure is not apparent [20,23]. This poses limitations on the use of such databases for patient care and research.

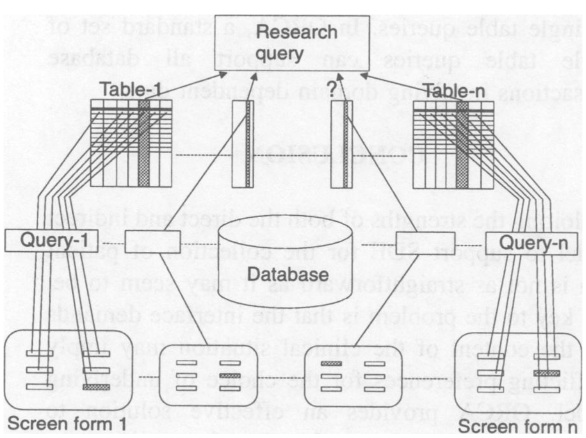

Fig. 2. Schematic representation of the direct model. One concept may be represented by more than one attribute.

Reliable retrieval requires semantic uniformity, which means that there is a one-to-one relationship between representation and meaning. Semantic uniformity in a direct model requires the use of views on a set of tables with no overlapping attributes or very large tables. The use of large tables for general use would easily lead to sparsely filled records. Indirect models have the advantage that they permit selective data entry without the problem of sparsely filled records.

Sharing of records. When records differ in content and organization, navigation in those records becomes more difficult for co-treating physicians and brings the risk of overlooking information. Physicians will benefit from a common vocabulary and a common interface when they share their records. This is best supported with an indirect model where the vocabulary takes care of a common terminology and the data structure of a common interface.

Consensus. A fixed-form interface is always a compromise for its users. An indirect model with a controlled vocabulary that covers a wide domain, has the advantage that physicians do not have to agree on the scope and detail of what they want to enter. As long as the appropriate items are available in the vocabulary, they are available for data entry.

THE CHALLENGE

In one clinical setting, the preferred interface strategy and data category may lead to conflicting preferences for the type of underlying datamodel. When the domain dependence of the data involved is decisive for the choice of model, two challenges arise:

1. How to provide an efficient interface for small well-defined, but domain dependent data sets?
2. How to support clinical settings involving both domain-dependent and domain-independent data?

RECONCILING MODEL TYPE AND INTERFACE DEMANDS

Custom form views on the indirect model

The first challenge is restricted to domain dependent data: can an indirect model be reconciled with a form-based interface? Although indirect models are most often used in combination with flexible knowledge-driven interfaces for data entry, they do not preclude the definition of specific views on the contents of the controlled vocabulary. In ORCA, the controlled vocabulary consists of a concept thesaurus and a concept network, that defines how the concepts may be combined into meaningful expressions. A parent and its children are represented by triplets of the form: parent - sequence number - child. In ORCA, the user can interactively define views on the contents of the network. A network browser enables the user to combine on one form the items that he wishes to describe in the context of a medical procedure.

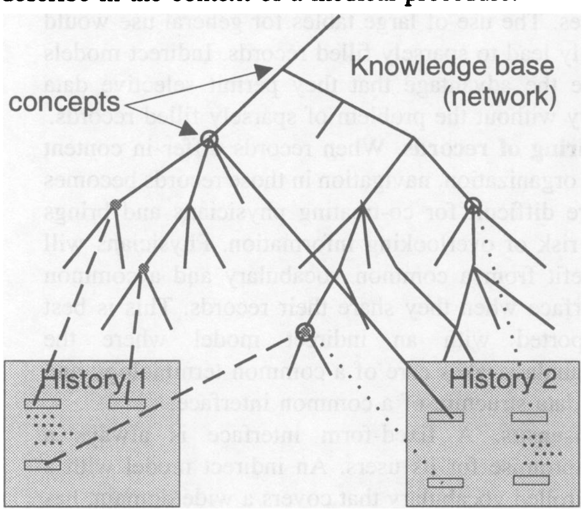

Fig. 3. Views map items on a form with a path in the knowledge base. Overlapping items correspond with instantiations of only one concept in the knowledge base.

Each selected item appears on the form with a link to that item in the network. This link takes the form of a sequence of items, that represents the path from the topnode of the network to the selected item.

Using paths eliminates ambiguity that could otherwise occur when one item can be described in multiple contexts. The user can arrange the items on the form according to his preferences and add labels to the form to organize the layout. The forms can be created at institutional, departmental, or personal level.

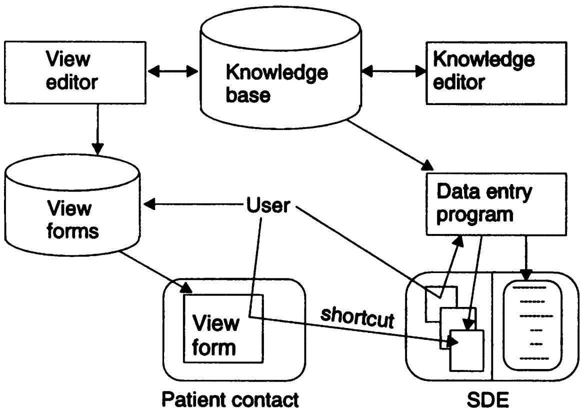

Fig. 4. Views can be defined interactively. The corresponding forms provide shortcuts for SDE.

During data entry, a pull-down menu allows selection of one of the available forms. The items on the selected form serve two purposes. In data entry mode, selection of an item activates the link to the concept network with the effect that the user can immediately proceed with his description of that item. In the background, the concept sequence defined by the link is instantiated automatically as if the user had made those selections himself. In consultation mode, selection of an item immediately highlights it among the entered data if present.

Views involving both models

The second challenge pertains to clinical settings involving a combination of domain dependent and domain independent data. A typical example are drug prescriptions during a patient visit. History and physical examination are typically domain-dependent, whereas a drug prescription is not. Since ORCA adopts the principle of semantic uniformity, each item can only be represented once. When domain dependence is decisive for the choice of model, a combined view is needed on parts of both models to support an interface that includes both data categories. Views on the controlled vocabulary pose no problems and function as described. Because of potential ambiguity and referential integrity, most relational DBMSs only support multiple table views for retrieval and not for insert and update transactions [24]. Due to the nature of the data represented in the direct model, this poses no problem in ORCA. For each domain dependent item (e.g. medication or lab test) in the data set, the descriptors will always correspond with the fields of one underlying table. As a consequence, there is no need for a variety of complex multiple table views that each need to be decomposed in a set of single table queries. In ORCA, a standard set of single table queries can support all database transactions involving domain dependent data.

CONCLUSION

Exploiting the strengths of both the direct and indirect model to support SDE for the collection of patient data is not as straightforward as it may seem to be. The key to the problem is that the interface demands and the content of the clinical situation may imply conflicting preferences for the choice of underlying model. ORCA provides an effective solution to reconcile the demands posed by interface and content by the way it adheres to the requirement of semantic uniformity. The essence is the distinction between domain dependent data and domain independent data, and the decision to support these categories with a direct model and an indirect model respectively.

Views on domain dependent data are supported via user-definable custom views, whereas the problem of many domain specific multiple table views is eliminated. Recording of every combination of domain independent data item can be supported by one or more standard single table queries.

With this strategy ORCA provides an important step forward towards a record that combines flexibility in content with the efficiency, structure, and semantic transparency to support (shared) patient care in a variety of domains, and the extraction of data for purposes that require standardized analysis.

In the context of the I4C project (Integration and Communication for the Continuity of Cardiac Care) of the European Fourth Framework Programme, ORCA is being installed at sites in six European countries.

References

1. McDonald CJ, Tierney WM. Computer-stored medical records: Their future role in Medical practice. JAMA 1988;259:3433-40
2. Reiser J. The Clinical Record in Medicine. Part 2: Reforming Content and Purpose. Ann Intern Med 1991;114:980-85.
3. Shortliffe EH, Tang PC. Patient records and computers. Ann Intern Med 1991;115:979-81
4. Dick RS, Steen EB. The computer-based patient record: an essential technology for health care. Committee on Improving the Patient record Division of Health Care Services. Institute of Medicine. National Academy press 1991.
5. Cimino JJ. Data storage and knowledge representation for clinical workstations. Int J Biomed Comput 1994;34:185-94.
6. Baud RH, Rassinoux AM, Scherrer JR. Natural Language Processing and Semantical Representation of Medical Texts. Meth Inform Med 1992;31:117-26.
7. Rasmussen JENG, Bassoe CF. Semantic analysis of medical records. Methods Inform Med 1993;32(1):66-72.
8. Barnett GO. The application of computer-based medical-record systems in ambulatory practice. N Engl J Med 1984;310:1643-50.
9. Kent DL, Shortliffe EH, Carlson RW, Bischoff MB, Jacobs CD. Improvements in data collection through physician use of a computer-based chemotherapy treatment consultant. J Clin Oncol 1985;3(10):1409-17.
10. Whiting-O'Keefe QE, Simborg DW, Epstein WV, Warger A. A computerized summary medical record system can provide more information than the standard medical record. Jama 1985;254(9):1185-92.
11. Wyatt JC. Clinical data systems, Part 3: Development and evaluation. Lancet 1994;344(8938):1682-8.
12. Howkins TJ, Kay S, Rector AL, et al. An overview of the PEN & PAD project. In: Rienhoff O, Lindberg DAB, eds. Lecture Notes in Medical Informatics 1990;40:73-78.
13. Trace D, Naeymi-Rad F, Haines D, et al. Intelligent Medical Record--entry (IMR-E). J Med Syst 1993;17(3-4):139-51.
14. Gouveia-Oliveira A, Salgado NC. A unified approach to the design of clinical reporting systems. Meth Inform Med 1994;33:479-87.
15. Bernauer J. Conceptual graphs as an operational model for descriptive findings. In: Clayton PD, ed. *Proceedings of the 15th SCAMC*. New York: McGraw-Hill 1991:214-8.
16. Poon AD, Fagan LM. The design and evaluation of a pen-based computer system for structured data entry. In: Ozbolt JG, ed. *Proceedings of the 18th SCAMC*. Special issue of JAMIA 1994:447-51.
17. Moorman PW, van Ginneken AM, Siersema PD, et al. Evaluation of Reporting Based on Descriptional Knowledge. JAMIA 1995;2:365-73.
18. van Ginneken AM, Stam H. Moorman PW. A multi-strategy approach for medical records of specialists. Int J Biomed Comput 1996;42:21-6.
19. van Ginneken AM. Structured data entry in ORCA: the strengths of two models combined. In: Cimino JJ, ed. J Am Med Inf Assoc, symposium supplement, 1996:797-801.
20. Pierik FH, van Ginneken AM, Timmers T, Stam H, Weber RFA. Restructuring routinely collected patient data: ORCA applied to andrology. Accepted for publication in Meth Inform med.
21. Cimino JJ, Clayton PD, Hripsack G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc 1994;1(1):35-50.
22. Evans A, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. J Am Med Informatics Assoc 1994;1(3):207-217.
23. Zwetsloot-Schonk JH, van Stiphout WA, Snitker P, et al. How to approach a hospital information system as a sampling frame: selection of patients with a percutaneous renal biopsy. Med Inf 1991;16:287-98.
24. Elmasri R, Navathe SB. Updating of Views and View Implementation. In: Fundamentals of Database Systems. The Benjamin/Cummings Publishing Company, New York. 1994:217-8.

E-mail first author: vanginneken@mi.fgg.eur.nl